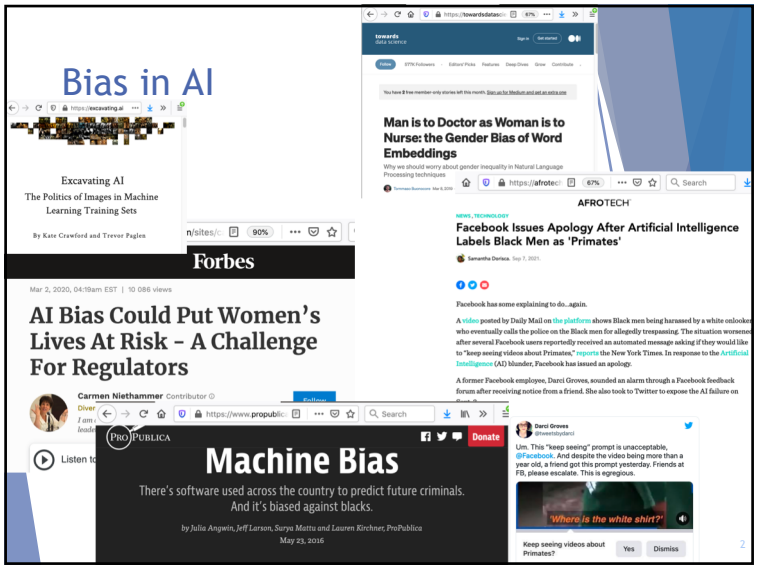
A preliminary assessment of bias in ontologies

Their sources and some consequences

C. Maria Keet
Department of Computer Science
University of Cape Town, South Africa

Talk at the Competence Center on Explainability, Fairness and Acceptability of Intelligent Systems (EFA), University of Ulm, Germany (online), 23 February 2022

Bias in AI



What about ontologies and knowledge graphs?

- Very little investigation into it
  - Friend of a Friend vocabulary [Gomes20]
  - Data aggregation for the Dirty War index [Keet09]
  - Exploratory notes on knowledge graphs [Janowicz17]
- Google's Knowledge Graph[1]:

[1] https://blog.google/products/search/about-knowledge-graph-and-knowledge-panels/
images: screengrabs from: https://www.youtube.com/watch?v=mmQl6VGvX-c

3

[1] https:
images: screengrabs from: https://www.youtube.com/watch?v=mmQl6VGvX-c

4

What about ontologies and knowledge graphs?

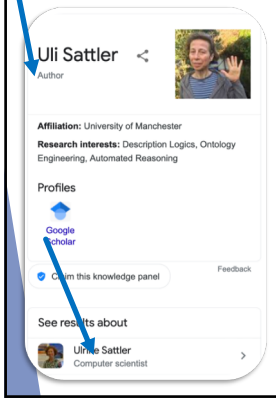
- Very little investigation into it
  - Friend of a Friend vocabulary [Gomes20]
  - Data aggregation for the Dirty War index [Keet09]
  - Exploratory notes on knowledge graphs [Janowicz17]
- Google's Knowledge Graph[1]:
  - The person who builds and controls the ontology or knowledge graph, then, is the one who has the power to control presentation and access to information and possibly also the recording of information.
  - "to some degree contests the autonomy of the user" [Vang13]

[1] https://blog.google/products/search/about-knowledge-graph-and-knowledge-panels/
images: screengrabs from: https://www.youtube.com/watch?v=mmQl6VGvX-c

5

Annotation and retrieval – Google's Knowledge Graph mess

German computer scientists…



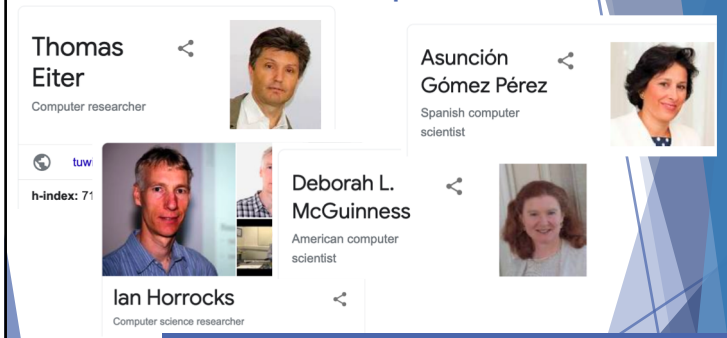
6

Annotation an… Google's Know…

German computer scientists…

7

More terms for computer scientists…



… For sure we won't all be easily *findable* with a simple query … nor do we have the authority to fix this mess … nor do we know all the terms to determine whether there is a bias in labeling

8

Aims

- Contribute to systematising the sort of bias that may be present in ontologies and similar artefacts
- Provide a preliminary answer to what bias means for ontologies, what their sources are, and how that manifests itself in ontologies
- Assess it across a set of ontologies in the same domain

Keet, C.M. An exploration into cognitive bias in ontologies. *Cognition And OntologieS (CAOS'21), part of JOWO'21*. 13-16 Sept 2021, Bolzano, Italy. Sanfilippo, E.M. et al. (Eds.). CEUR-WS vol. 2969. 17p

9

This talk

- Some preliminary considerations
- Identify, discuss, illustrate sources of bias
- Evaluation: assess three COVID-19 ontologies
- Automated reasoning considerations

10

Principal sources of bias in ontologies

11

Preliminaries

- Defining cognitive bias… and differentiate from cognitive styles, alternate perspectives, image schemas, simple mistakes
- Inclusive definition for bias is adopted:
  - "a consequence of interference with honest attempts" [Oreg, 2009]
- Variants: narrow scope of *norm deviation* and *error*
- Implicit vs explicit
- For IT and computing, grouped by dimension; e.g.,
  - by type of task for information visualisation [Dimara20]
  - by software engineering "knowledge area" [Mohanani20]

12

Sampling of cognitive biases from Dimara et al.'s list

- 17 of the 154 seem potentially applicable to ontologies; e.g.:
  - **Mere exposure/familiarity**: choice is influenced by exposure to it and thus familiarity with it.
  - **Naive realism**: the belief that you experience objects in your world objectively.
  - **False Consensus**: Overestimating that other people are and behave like you and agree with your opinion.
  - **Barnum effect**: High accuracy ratings for vague and general statements. (?)

13

Possible biases, by source

Summary of typical possible biases in ontologies grouped by source

Type	Subtype	[im/ex]plicit bias
Philosophical	-	explicit
Purpose	-	explicit
Subject domain	Science	explicit
	Granularity	either
	Linguistic	either
	Socio-cultural	either
	Political or religious	either
	Economics	explicit

14

Foundational ontology differences (philosophical or otherwise)

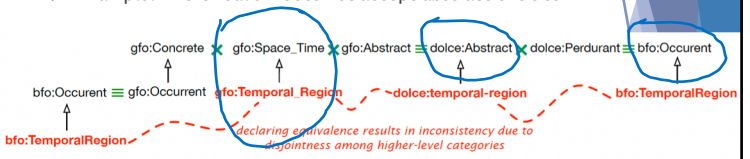
- Realism vs idealism, concepts, universals etc.
- Some differences don't matter in praxis; some do
- Ways to find and resolve the (explicit!) conflict(s)
- Example: BFO's realism does not accept abstract entities

Keet, C.M., Grütter, R. Toward a systematic conflict resolution framework for ontologies. *Journal of Biomedical Semantics*, 2021, 12:15.

15

Foundational ontology differences (philosophical or otherwise)

- Realism vs idealism, concepts, universals etc.
- Some differences don't matter in praxis; some do
- Ways to find and resolve the (explicit!) conflict(s)
- Example: BFO's realism does not accept abstract entities



Khan, Z.C., Keet, C.M. Foundational ontology mediation in ROMULUS. *Knowledge Discovery, Knowledge Engineering and Knowledge Management: IC3K 2013 Selected Papers*. A. Fred et al. (Eds.). Springer CCIS vol. 454, pp. 132-152, 2015.

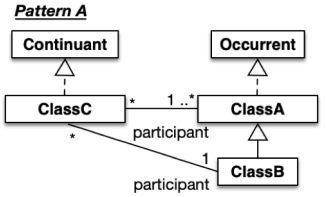
Keet, C.M., Grütter, R. Toward a systematic conflict resolution framework for ontologies. *Journal of Biomedical Semantics*, 2021, 12:15.

16

Purpose: *encoding* bias rather than a cognitive bias

Three different patterns with a purpose bias:

- Pattern A: biased toward a scientific approach
- Pattern B: conceptual data modelling influence or purpose
- Pattern C: a thesaurus-like approach useful for document annotation



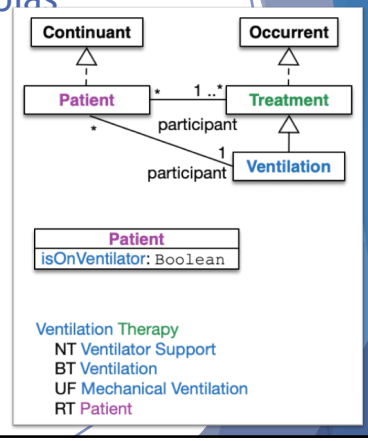
Fillottrani, P.R., Keet, C.M.. Dimensions Affecting Representation Styles in Ontologies. *1st Iberoamerican conference on Knowledge Graphs and Semantic Web (KGSWC'19)*. Springer CCIS vol. 1029, 186-200.

17

Purpose: *encoding* bias rather than a cognitive bias

Three different patterns with a purpose bias:

- Pattern A: biased toward a scientific approach
- Pattern B: conceptual data modelling influence or purpose
- Pattern C: a thesaurus-like approach useful for document annotation



Fillottrani, P.R., Keet, C.M.. Dimensions Affecting Representation Styles in Ontologies. *1st Iberoamerican conference on Knowledge Graphs and Semantic Web (KGSWC'19)*. Springer CCIS vol. 1029, 186-200.

18

Granularity

- How detailed should it be? How many levels in the taxonomy?
- Less details…
  - Act of omission
    - E.g.: aggregating ex-military persons with non-involved persons (civilians) as one group of 'Civilians'
  - Not needed
  - Ran out of time
- … but may be very difficult to determine, unless stated
- Not needed and ran out of time may be explicit choices and prioritization (explicit bias), or honestly out of scope for v1 or ever

19

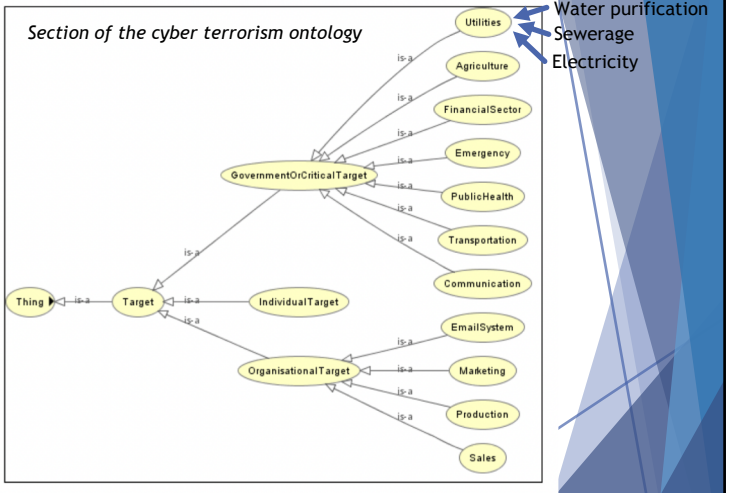


Figure 4: Target Class [Veerasamy12]

20

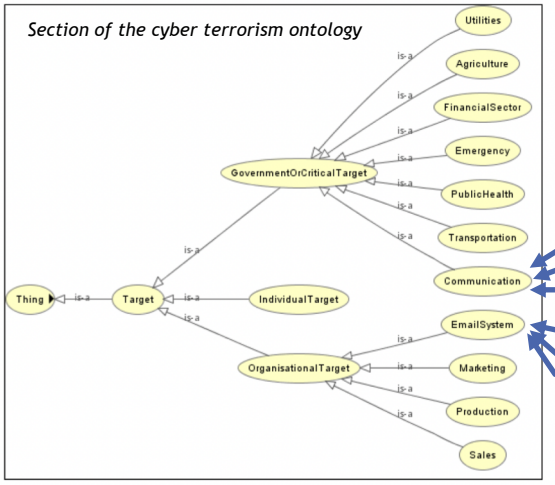


Figure 4: Target Class [Veerasamy12]

21

Socio-cultural issues in ontologies

- Relates to how society is organised, the assumptions that underlie it and history
  - organisational structures, who lives with whom, demographics, allocation of resources, or social geography
- Example: GoodRelations Ontology's payment methods and legal status of a Business [Hepp08]

https://shop.ikhokha.com/collections/card-machines/products/moyer-pro
"Promoting Work and Networks for Women in Informal Economy- A Demonstration of Proactive Interventions, New Delhi" by [redacted] is licensed under CC BY-NC-ND 2.0
"Geoffrey's Malibu Sunday Brunch libations" by Al_HikesAZ is licensed under CC BY-NC 2.0

22

Socio-cultural issues in ontologies

- Relates to how society is organised, the assumptions that underlie it and history
  - organisational structures, who lives with whom, demographics, allocation of resources, or social geography
- Example: GoodRelations Ontology's payment methods and legal status of a Business [Hepp08]
- Alcohol use and disorder across time and cultures
  - DSM-IV, DSM-V, and ICD-10 coding differences
  - modifications due to a combination of socio-cultural factors and some scientific disagreement [Wakefield15]

https://shop.ikhokha.com/collections/card-machines/products/moyer-pro
"Promoting Work and Networks for Women in Informal Economy- A Demonstration of Proactive Interventions, New Delhi" by [redacted] is licensed under CC BY-NC-ND 2.0
"Geoffrey's Malibu Sunday Brunch libations" by Al_HikesAZ is licensed under CC BY-NC 2.0

23

Political motivations

- The easy one: Aggrieved group vs Terrorist organisation
- Concretely,
  - *terrorist* and *terroristgroup* in the terrorism ontology [Jindal20]
  - *ActorEntity* with various types of *Insiders* and *Protestors* in the Cyberterrorism ontology [Veerasamy12]
- CIDO's "Wuhan virus"

"Mural, IRA-Wandbild in Belfast, women" by PPCC Antifa is licensed under CC BY-NC-SA 2.0

24

Evaluation: the COVID-19 ontologies

25

COVID-19 ontologies

- 'same' topic on COVID-19, developed at same time by different groups
  - Coronavirus Infectious Disease Ontology (CIDO) [He20]
  - COviD-19 Ontology (CODO) [Dutta20]
  - Coronavirus Vocabulary (COVoc) [Pendlington20]
- Assess their documentation, characteristics, content
- Iteratively note observations of bias and check subset of cognitive bias list and consider wrt the ontologies

26

Bias by cognitive bias

Bias *(Cognitive biases from Dimara et al's list)*	**CIDO**	**CODO**	**COVoc**
Mere exposure/familiarity (choice is influenced by exposure to it and thus familiarity with it)	+		+
Negative interpretation (judgement is affected more by negative information than positive)	+		
Optimism (more positive predictions for oneself than for others)	+		
Naive realism (the belief that you experience objects in your world objectively)	+		
False Consensus (Overestimating that other people are and behave like you and agree with your opinion)		+	
Illusory truth effect (a statement is considered to be true after repeated exposure to it)			+

27

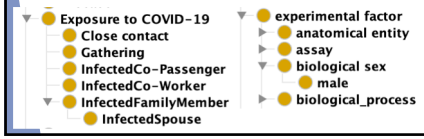
Presence/absence, by source

Bias *(Source/type)*	**CIDO**	**CODO**	**COVoc**
Philosophical	+	−	+
Purpose	−	+	+
Science	−	−	+
Granularity	±	+	+
Linguistic	+	−	−
Socio-cultural	+	+	+
Political or religious	+	+	+
Economics	−	−	±

28

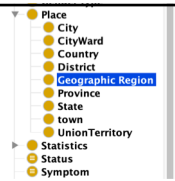
Presence/absence, by source

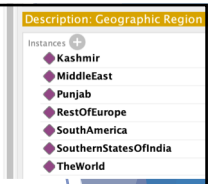
Bias *(Source/type)*	CIDO	CODO	COVoc
Philosophical	+	−	+
Purpose	−	+	+
Science	−	−	+
Granularity	±	+	+
Linguistic	+	−	−
Socio-cultural	+	+	+
Political or religious	+	+	+
Economics	−	−	±



29

Presence/absence,





Bias *(Source/type)*	CIDO	CODO	COVoc
Philosophical	+	−	+
Purpose	−	+	+
Science	−	−	+
Granularity	±	+	+
Linguistic	+	−	−
Socio-cultural	+	+	+
Political or religious	+	+	+
Economics	−	−	±

30

Automated reasoning

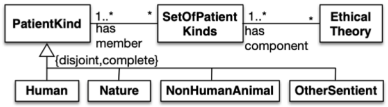
31

Any effects on automated reasoning?

- 'Incoherence' (one or more unsatisfiable classes), inconsistencies, or undesirable deductions
- For TBox only: incoherence and undesirable deductions will be found at authoring time already, not during deployment
- For the knowledge base (Tbox + Abox [in owl or secondary storage]): inconsistencies or undesirable deductions either at authoring time or during deployment
  - Querying data
  - Annotating data (and subsequent retrieval)

32

Domain ontology example – desirable or undesirable deduction

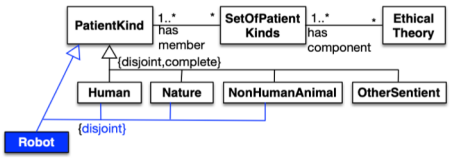


Section of the Genet.owl ontology:
PatientKind ≡ Human ⊔ Nature ⊔ NonHumanAnimal ⊔ OtherSentient
Human ⊓ Nature ⊑ ⊥ (etc. for each pair of classes)

Rautenbach, J.G., Keet, C.M. Toward equipping Artificial Moral Agents with multiple ethical theories. *RobOntics: International Workshop on Ontologies for Autonomous Robotics*, co-located with BoSK'20, Bolzano. CEUR-WS vol. 2708, 5

33

Domain ontology example – desirable or undesirable deduction

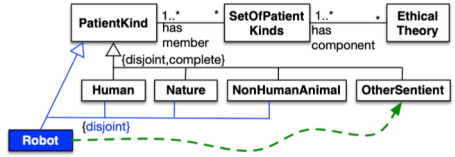


Section of the Genet.owl ontology:
PatientKind ≡ Human ⊔ Nature ⊔ NonHumanAnimal ⊔ OtherSentient
Human ⊓ Nature ⊑ ⊥ (etc. for each pair of classes)

Disagree and add to the ontology:
Robot ⊑ PatientKind
Robot ⊓ Human ⊑ ⊥
Robot ⊓ Nature ⊑ ⊥
Robot ⊓ NonHumanAnimal ⊑ ⊥

34

Domain ontology example – desirable or undesirable deduction



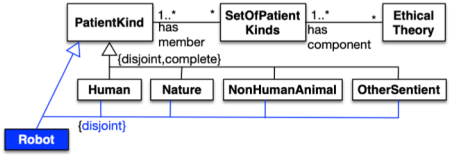
Section of the Genet.owl ontology:
PatientKind ≡ Human ⊔ Nature ⊔ NonHumanAnimal ⊔ OtherSentient
Human ⊓ Nature ⊑ ⊥ (etc. for each pair of classes)

Disagree and add to the ontology:
Robot ⊑ PatientKind
Robot ⊓ Human ⊑ ⊥
Robot ⊓ Nature ⊑ ⊥
Robot ⊓ NonHumanAnimal ⊑ ⊥

Deduction:
Robot ⊑ OtherSentient

35

Domain ontology example – desirable or undesirable deduction

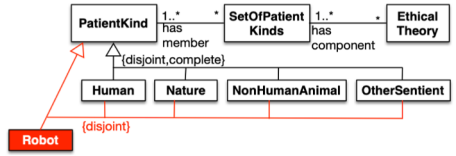


Section of the Genet.owl ontology:
PatientKind ≡ Human ⊔ Nature ⊔ NonHumanAnimal ⊔ OtherSentient
Human ⊓ Nature ⊑ ⊥ (etc. for each pair of classes)

Disagree and add to the ontology:
Robot ⊓ OtherSentient ⊑ ⊥

36

Domain ontology example – desirable or undesirable deduction



Section of the Genet.owl ontology:

PatientKind ≡ Human ⊔ Nature ⊔ NonHumanAnimal ⊔ OtherSentient

Human ⊓ Nature ⊑ ⊥ (etc. for each pair of classes)

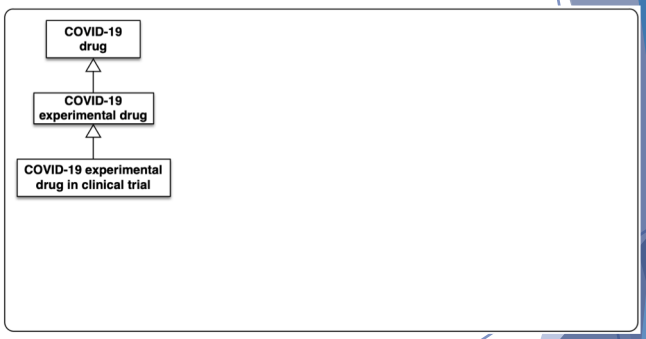
Disagree and add to the ontology:

Robot ⊓ OtherSentient ⊑ ⊥

Deduction:

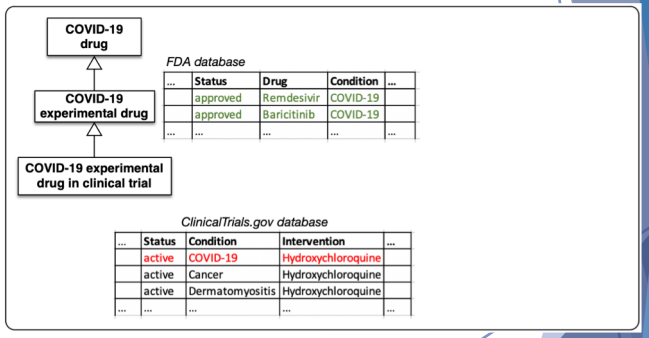
Robot ⊑ ⊥

OBDA toy data integration example with the CIDO

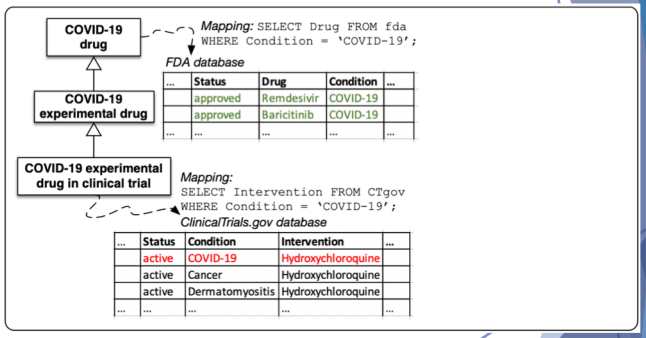


Optimism bias

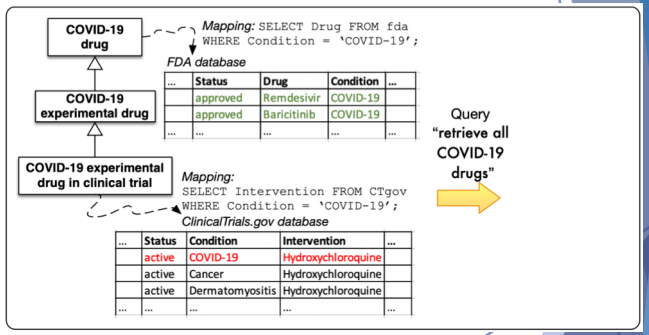
OBDA toy data integration example with the CIDO



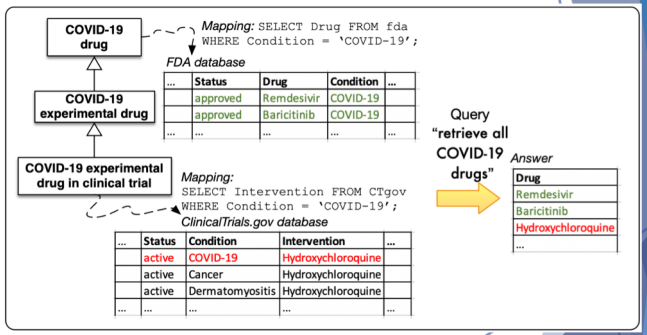
OBDA toy data integration example with the CIDO



OBDA toy data integration example with the CIDO



OBDA toy data integration example with the CIDO



Discussion

- Ontological investigation of bias?
- Can an ontology ever be free of bias?
- That loose end on bias with "honest attempts" vs modelling mistake vs 'ran out of time':
  - Need a way to disambiguate
  - How can one be certain it is a bias when not involved in the development of that ontology? (but if one is, one may be blind to the bias)
- Consequences for automated reasoning

Conclusions

- Bias may be present in an ontology, a number of which can be categorised as cognitive biases
- Eight categories of sources of bias for ontologies: philosophical, purpose, science, granularity, linguistic, socio-cultural, political or religious, and economic motives
- Three COVID-19 ontologies each exhibited a different subset of the sources of bias and cognitive biases
- This first characterisation and comparative assessment may contribute to further research into cognitive bias of ontologies, its methods, and definitions

References

[Dimara20] E. Dimara, S. Franconeri, C. Plaisant, A. Bezerianos, P. Dragicevic, A task-based taxonomy of cognitive biases for information visualization, IEEE Transactions on Visualization and Computer Graphics 26 (2020) 1413-1432

[Dutta20] B. Dutta, M. DeBellis, CODO: an ontology for collection and analysis of COVID-19 data, in: IC3K 2020, INSTICC, 2020.

[He20] Y.He,H.Yu,E.Ong,Y.Wang,Y.Liu,A.Huffman,H.huiHuang,J.Beverley,A.Y.Lin,W.D. Duncan, S. Arabandi, J. Xie, J. Hur, X. Yang, L. Chen, G. S. Omenn, B. Athey, B. Smith, Cido: The community-based coronavirus infectious disease ontology, in: ICBO'20, volume 2807, CEUR-WS, 2020.

[Hepp08] M. Hepp, Goodrelations: An ontology for describing products and services offers on the web, in: EKAW'08, volume 5268 of *LNCS*, Springer, 2008, pp. 332-347.

[Gomes20]] DL Gomes, TH.Bragato Barros, The bias in ontologies: An analysis of the FOAF ontology. Proc of ISKO'20.

[Janowicz18] K Janowicz, B Yan, B Regalia, R Zhu, G Mai, Debiasing knowledge graphs: Why female presidents are not like female popes. Proc. of ISWC'18 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks.

[Jindal20] R. Jindal, K. Seeja, S. Jain, Construction of domain ontology utilizing formal concept analysis and social media analytics, International Journal of Cognitive Computing in Engineering 1 (2020) 62 - 69.

[Keet09] CM Keet. Dirty wars, databases, and indices, Peace & Conflict Review 4 (2009) 75-78.

[Mohanani20] R. Mohanani, I. Salman, B. Turhan, P. Rodríguez, P. Ralph, Cognitive biases in software engineering: A systematic mapping study, IEEE Transactions on Software Engineering 46 (2020) 1318-1339.

[Oreg, 2009] S. Oreg, M. Bayazit, Prone to bias: Development of a bias taxonomy from an individual differences perspective, Review of General Psychology 13 (2009) 175-193.

[Pendlington20] Z. M. Pendlington, P. Roncaglia, N. Matentzoglu, D. Osumi-Sutherland, D. Caucheteur, J. Gobeill, L. Mottin, D. Agosti, P. Ruch, H. Parkinson, COVoc: a COVID-19 ontology to support literature triage, 2020. URL: https://raw.githubusercontent.com/CIDO-ontology/ WCO/master/day-1/Zoe_COVoc.pdf, WCO-2020: Workshop on COVID-19 Ontologies.

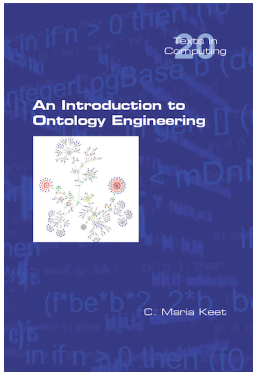
[Vang13] K Vang, Ethics of Google's Knowledge Graph: some considerations, J. Inf. Comm. & Ethics in Society 11 (2013) 245-260.

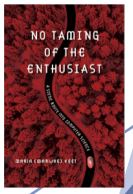
[Veerasamy12] N. Veerasamy, M. Grobler, B. V. Solms, Building an ontology for cyberterrorism, in: E. Filiol, R. Erra (Eds.), Proc. 11th European Conference on Information Warfare and Security, Academic Publishing International, 2012, pp. 286-295.

[Wakefield15] J. C. Wakefield, DSM-5 substance use disorder: How conceptual missteps weakened the foundations of the addictive disorders field, Acta Psychiatrica Scandinavica 132 (2015) 327-334.

A bit of advertisement

https://people.cs.uct.ac.za/~mkeet/OEbook/





http://www.meteck.org/ntote

Thank you for your attention

Questions?

Email: mkeet@cs.uct.ac.za

CS Dept & School of IT: http://www.sit.uct.ac.za/

Homepage: http://www.meteck.org/

Blog: http://keet.wordpress.com